

A note on Maximum Likelihood Estimation for cubic and quartic canonical toric del Pezzo Surfaces

Dimitra Kosta

February 29, 2016

Abstract

In this article a closed-form for the Maximum Likelihood Estimate of algebraic statistical models which correspond to cubic and quartic toric del Pezzo surfaces with Du Val singular points is given.

1 Introduction

Maximum likelihood estimation (MLE) is a standard approach to parameter estimation and inference, and a fundamental computational task in statistics. It consists of the following problem: given the observed data and a model of interest, find the probability distribution that is most likely to have produced the data. In the past decade, algebraic techniques for the computation of maximum likelihood estimates have been developed with some success for algebraic statistical models for discrete data (see [1], [3], [4]).

This article is concerned with the problem of Maximum Likelihood Estimation for algebraic statistical models with singularities, in particular those which correspond to toric del Pezzo surfaces with Du Val singular points. The importance of algebraic statistical models corresponding to toric varieties is due to their relation to log-linear statistical models which are widely used in Statistics ([2]). Another reason for studying the MLE for such algebraic statistical models, is that they correspond to singular varieties. Singularities play an important role in statistical inference as the commonly assumed smoothness of algebraic statistical models is very restrictive and is almost never satisfied for models of statistical relevance (see [[2], p. 100], [8]).

The relevant definitions to this problem are given bellow.

The author would like to thank Ivan Cheltsov, Alexander Davie and Milena Hering for valuable comments and corrections.

1.1 Maximum Likelihood Estimation

Consider the complex projective space \mathbb{P}^n with coordinates (p_0, p_1, \dots, p_n) . In our setting the coordinate p_i represents the probability of the i -event therefore $p_0 + p_1 + \dots + p_n = 1$. The set of points in \mathbb{P}^n with positive real coefficients is identified with the probability simplex

$$\Delta_n = \{(p_0, p_1, \dots, p_n) \in \mathbb{R}^{n+1} : p_0, p_1, \dots, p_n \geq 0 \text{ and } p_0 + p_1 + \dots + p_n = 1\}.$$

An algebraic statistical model is a closed subset \mathcal{M} of the complex projective space \mathbb{P}^n , with the model itself being the intersection of \mathcal{M} with the probability simplex Δ_n . The data is given by a non-negative integer vector $(u_0, u_1, \dots, u_n) \in \mathbb{N}^{n+1}$, where u_i is the number of times the i -event is observed.

The maximum likelihood estimation problem consists of finding a model point p which maximises the likelihood of observing the data. This amounts to maximising the corresponding Likelihood Function

$$L(p_0, p_1, \dots, p_n) = \frac{p_0^{u_0} \cdot p_1^{u_1} \cdot \dots \cdot p_n^{u_n}}{(p_0 + p_1 + \dots + p_n)^{(u_0 + u_1 + \dots + u_n)}}$$

over the model \mathcal{M} , where here we ignore a multinomial coefficient. Statistical computations are usually implemented in the affine n -plane $p_0 + p_1 + \dots + p_n = 1$. However, including the denominator makes the likelihood function a well-defined rational function on the projective space \mathbb{P}^n , enabling one to use projective algebraic geometry to study its restriction to the variety \mathcal{M} .

The likelihood function might not be convex, so it can have many local maxima and the problem of finding and certifying a global maximum is difficult. Therefore, in most recent works the problem of finding all critical points of the likelihood function is considered, with the aim of identifying all local maxima (see [1], [3] and [4]). This corresponds to solving a system of polynomial equations and these equations, defining the critical points of the likelihood function L , are called likelihood equations. The number of complex solutions to the likelihood equations equals the number of complex critical points of the restriction of the likelihood function L to the model \mathcal{M} , which is called the Maximum Likelihood (ML) degree of the variety \mathcal{M} .

1.2 Toric models

In this article we are studying the Maximum Likelihood Estimation problem for toric models which are models with a well behaved likelihood function. Toric models are known as log-linear models in statistics, because the logarithms of the probabilities are linear functions in the logarithms of the parameters θ_i . They have the property that maximum likelihood estimation is a convex optimization problem. Assuming that the parameter domain Θ is bounded, it follows that the likelihood function has exactly one local maximum. We introduce toric models following the notation used in Chapter 1.2 of [6].

Let $A = (a_{ij})$ be a non-negative integer $d \times m$ matrix with the property that all column sums are equal:

$$\sum_{i=1}^d a_{i1} = \sum_{i=1}^d a_{i2} = \dots = \sum_{i=1}^d a_{im} .$$

The j -th column vector a_j of the matrix A represents the monomial

$$\theta^{a_j} := \theta_1^{a_{1j}} \cdot \theta_2^{a_{2j}} \cdot \dots \cdot \theta_d^{a_{dj}} \text{ for all } j = 1, \dots, m.$$

and the assumption that the column sums of the matrix A are all equal means these monomials all have the same degree.

Definition 1.1 *The toric model of A is the image of the orthant $\Theta = \mathbb{R}_{>0}^d$ under the map*

$$f : \mathbb{R}^d \rightarrow \mathbb{R}^m, \theta \mapsto \frac{1}{\sum_{j=1}^m \theta^{a_j}} \cdot (\theta^{a_1}, \theta^{a_2}, \dots, \theta^{a_m}) .$$

Maximum likelihood estimation for the toric model means solving the optimization problem of maximising the function

$$p_1^{u_1} \cdot p_1^{u_2} \cdot \dots \cdot p_m^{u_m}$$

subject to the constraints $f(\mathbb{R}_{>0}^d)$. This is equivalent to maximising function

$$\theta^{A \cdot u} \text{ subject to } \theta \in \mathbb{R}_{>0}^d \text{ and } \sum_{j=1}^m \theta^{a_j} = 1,$$

where

$$\theta^{Au} = \prod_{i=1}^d \theta_i^{a_{i1}u_1 + a_{i2}u_2 + \dots + a_{im}u_m} \text{ and } \theta^{a_j} = \prod_{i=1}^d \theta_i^{a_{ij}} .$$

Let $b := Au$ denote the sufficient statistic, then the optimisation problem above becomes

$$\text{Maximise } \theta^b \text{ subject to } \theta \in \mathbb{R}_{>0}^d \text{ and } \sum_{j=1}^m \theta^{a_j} = 1 .$$

Proposition 1.2 *Fix a toric model A and data $u \in \mathbb{N}^m$ with sample size $N = u_1 + \dots + u_m$ and sufficient statistic $b = Au$. Let $\hat{p} = f(\hat{\theta})$ be any local maximum for the equivalent optimization problems above. Then*

$$A \cdot \hat{p} = \frac{1}{N} \cdot b.$$

Given a matrix $A \in \mathbb{N}^{d \times m}$ and any vector $b \in \mathbb{R}^d$, we consider the set

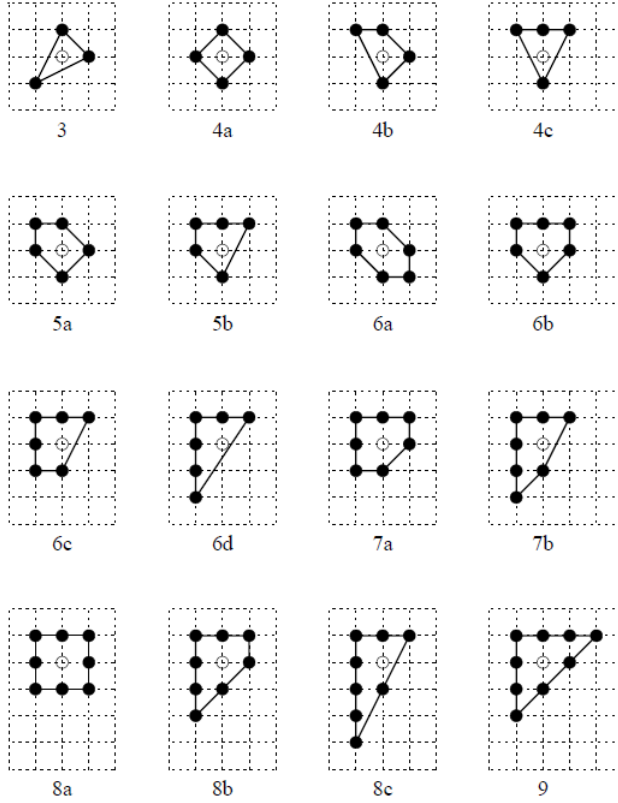
$$P_A(b) = \{p \in \mathbb{R}^m : A \cdot p = \frac{1}{N} \cdot b \text{ and } p_j > 0 \text{ for all } j\}.$$

This is a relatively open polytope and Birch's theorem below asserts that it is either empty or meets the toric model in precisely one point.

Theorem 1.3 (Birch's Theorem) *Fix a toric model A and let $u \in \mathbb{N}_{>0}^m$ be a strictly positive data vector with sufficient statistic $b = Au$. The intersection of the polytope $P_A(b)$ with the toric model $f(\mathbb{R}_{>0}^d)$ consists of precisely one point. That point is the maximum likelihood estimate \hat{p} for the data u .*

In the next section we will determine a closed form for the Maximum Likelihood Estimates for all algebraic statistical models corresponding to toric del Pezzo surfaces of degree three and four with Du Val singularities. The corresponding polytope of a toric del Pezzo surface with Du Val singularities is a reflexive polytope. According to the classification results of reflexive polytopes (see [5], [7]), there are 16 isomorphism classes of such two-dimensional reflexive polytopes, a list of which we provide bellow.

Proposition 1.4 *There are exactly 16 isomorphism classes of two-dimensional reflexive polytopes given in the list bellow. The number in the labels is the number of lattice points on the boundary.*

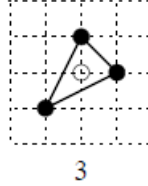


2 Main Calculation

In this section we determine a closed form for the Maximum Likelihood estimates for all algebraic statistical models corresponding to cubic and quartic toric del Pezzo surfaces with Du Val singularities.

2.1 Cubic del Pezzo with three singular points of type \mathbb{A}_2

Consider the case of a reflexive polytope with three lattice points on the boundary, as in the graph below. This corresponds to a cubic surface with three Du Val singular points of type \mathbb{A}_2 .



This polytope generated by the lattice points $(1,0), (0,1), (-1,-1)$ in \mathbb{Z}^2 gives the same toric variety as the polytope which is generated by the lattice points $(2,1,0), (1,2,0), (0,0,3)$ in \mathbb{Z}^3 .

We are interested in the algebraic statistical model given by the matrix

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}.$$

This non-negative integer matrix A corresponds to a toric algebraic variety which is the topological closure $\overline{f(\mathbb{C}^3)}$ of the image $f(\mathbb{C}^3) \subset \mathbb{C}^3$ under the map

$$f : \mathbb{C}^3 \rightarrow \mathbb{C}^3, (\theta_1, \theta_2, \theta_3) \mapsto \frac{1}{(\theta_1^2 \theta_2 + \theta_1 \theta_2^2 + \theta_3^3)} (\theta_1^2 \theta_2, \theta_1 \theta_2^2, \theta_3^3).$$

According to Birch's theorem there is a unique maximum likelihood estimate \hat{p} for the data $u = (u_1, u_2, u_3)$ with $N = u_1 + u_2 + u_3$. This unique MLE satisfies the equation

$$A \cdot \hat{p} = \frac{1}{N} \cdot A \cdot u.$$

Since $\det(A) \neq 0$, the inverse A^{-1} exists and by multiplying the above matrix equation with A^{-1} from the left we get

$$\hat{p} = \frac{1}{N} \cdot u.$$

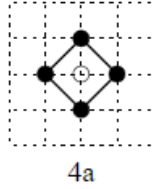
This gives us the equations

$$\begin{aligned}\hat{\theta}_1^2 \hat{\theta}_2 &= \frac{1}{N} u_1, \\ \hat{\theta}_1 \hat{\theta}_2^2 &= \frac{1}{N} u_2, \\ \hat{\theta}_3^3 &= \frac{1}{N} u_3.\end{aligned}$$

and we can compute $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3) = (\sqrt[3]{\frac{u_1}{N \cdot u_2}}, \sqrt[3]{\frac{u_2}{N \cdot u_1}}, \sqrt[3]{\frac{u_3}{N}})$.

2.2 Quartic del Pezzo with four singular points of type \mathbb{A}_1 .

Again consider the case of a reflexive polytope with four lattice points on the boundary, as in the graph below.



Again the polytope generated by the lattice points $(1, 0), (0, 1), (-1, 0), (0, -1)$ in \mathbb{Z}^2 gives rise to the same toric variety as the polytope generated by the lattice points $(2, 1, 0), (1, 2, 0), (1, 0, 2), (0, 1, 2)$ in \mathbb{Z}^3 .

The non-negative integer matrix

$$A = \begin{bmatrix} 2 & 1 & 1 & 0 \\ 1 & 2 & 0 & 1 \\ 0 & 0 & 2 & 2 \end{bmatrix}$$

corresponds to a toric algebraic variety which is the topological closure $\overline{f(\mathbb{C}^3)}$ of the image $f(\mathbb{C}^3) \subset \mathbb{C}^4$ under the map

$$f : \mathbb{C}^3 \rightarrow \mathbb{C}^4, (\theta_1, \theta_2, \theta_3) \mapsto \frac{1}{(\theta_1^2 \theta_2 + \theta_1 \theta_2^2 + \theta_1 \theta_3^2 + \theta_2 \theta_3^2)} (\theta_1^2 \theta_2, \theta_1 \theta_2^2, \theta_1 \theta_3^2, \theta_2 \theta_3^2).$$

According to Birch's theorem there is a unique maximum likelihood estimate $\hat{\theta}$ for the data $u = (u_1, u_2, u_3, u_4)$ which satisfies the equation

$$A \cdot \hat{p} = \frac{1}{N} \cdot A \cdot u,$$

where $N = u_1 + u_2 + u_3 + u_4$ and \hat{p} is the probability distribution corresponding to the parameter values $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$. This gives us the equations

$$\begin{aligned}\hat{\theta}_1^2 \hat{\theta}_2 - \hat{\theta}_2 \hat{\theta}_3^2 &= \frac{1}{N}(u_1 - u_4) \\ \hat{\theta}_1 \hat{\theta}_2^2 + \hat{\theta}_2 \hat{\theta}_3^2 &= \frac{1}{N}(u_2 + u_4) \\ \hat{\theta}_1 \hat{\theta}_3^2 + \hat{\theta}_2 \hat{\theta}_3^2 &= \frac{1}{N}(u_3 + u_4) ,\end{aligned}$$

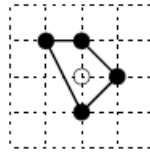
and these equations give us

$$\begin{aligned}\hat{p}_1 &= \frac{(u_1 + u_2)(u_1 + u_3)}{N^2} \\ \hat{p}_2 &= \frac{(u_1 + u_2)(u_2 + u_4)}{N^2} \\ \hat{p}_3 &= \frac{(u_1 + u_3)(u_3 + u_4)}{N^2} \\ \hat{p}_4 &= \frac{(u_2 + u_4)(u_3 + u_4)}{N^2} .\end{aligned}$$

We can compute the unique maximum likelihood estimate $\hat{\theta}$ for the data u , which is $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3) = (\sqrt[3]{\frac{(u_1+u_2)(u_1+u_3)^2}{N^2(u_2+u_4)}}, \sqrt[3]{\frac{(u_1+u_2)(u_2+u_4)^2}{N^2(u_1+u_3)}}, \sqrt[6]{\frac{(u_1+u_3)(u_2+u_4)(u_3+u_4)^3}{N^4(u_1+u_2)}})$.

2.3 Quartic del Pezzo with one \mathbb{A}_2 and two \mathbb{A}_1 type singular points

Again consider the case of a reflexive polytope with four lattice points on the boundary, as in the graph below.



4b

The polytope generated by the lattice points $(1, 0), (0, 1), (-1, 1), (0, -1)$ in \mathbb{Z}^2 is the same as the polytope generated by the lattice points $(2, 1, 0), (1, 2, 0), (1, 0, 2), (0, 2, 1)$ in \mathbb{Z}^3 .

The non-negative integer matrix

$$A = \begin{bmatrix} 1 & 0 & 1 & 2 \\ 0 & 2 & 2 & 1 \\ 2 & 1 & 0 & 0 \end{bmatrix}$$

corresponds to a toric algebraic variety which is the topological closure $\overline{f(\mathbb{C}^3)}$ of the image $f(\mathbb{C}^3) \subset \mathbb{C}^4$ under the map

$$f : \mathbb{C}^3 \rightarrow \mathbb{C}^4, (\theta_1, \theta_2, \theta_3) \mapsto \frac{1}{(\theta_1\theta_3^2 + \theta_2^2\theta_3 + \theta_1\theta_2^2 + \theta_1^2\theta_2)}(\theta_1\theta_3^2, \theta_2^2\theta_3, \theta_1\theta_2^2, \theta_1^2\theta_2) .$$

According to Birch's theorem there is a unique maximum likelihood estimate $\hat{\theta}$ for the data $u = (u_1, u_2, u_3, u_4)$ which satisfies the equation

$$A \cdot \hat{p} = \frac{1}{N} \cdot A \cdot u ,$$

where $N = u_1 + u_2 + u_3 + u_4$ and \hat{p} is the probability distribution corresponding to the parameter values $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$.

This gives us the equations

$$\begin{aligned} 2\hat{\theta}_1\hat{\theta}_3^2 + \hat{\theta}_1^2\hat{\theta}_2 &= \frac{1}{N}(2u_1 + u_4) \\ \hat{\theta}_2^2\hat{\theta}_3 - \hat{\theta}_1^2\hat{\theta}_2 &= \frac{1}{N}(u_2 - u_4) \\ 2\hat{\theta}_1\hat{\theta}_2^2 + 3\hat{\theta}_1^2\hat{\theta}_2 &= \frac{1}{N}(2u_3 + 3u_4) . \end{aligned}$$

Then the unique maximum likelihood estimate $\hat{\theta}$ for the data u is $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3) = (\sqrt[3]{\frac{p_4^2}{p_3}}, \sqrt[3]{\frac{p_3^2}{p_4}}, \sqrt[6]{\frac{p_3}{3p_4^2}[p_3 + \frac{3u_1 - u_3}{N}]})$. The probability distribution \hat{p}_3 and \hat{p}_4 are given by the quartic equations

$$a_i p_i^4 + b_i p_i^3 + c_i p_i^2 + d_i p_i + e_i = 0 \text{ for } i = 3, 4$$

where

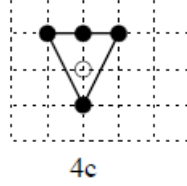
$$\begin{aligned} a_4 &= 19N^4 \\ b_4 &= 2N^3(19(u_2 - u_4) - 27N) \\ c_4 &= 9N^2(2u_3 + 3u_4)(6N - 6(u_2 - u_4) - 2(2u_3 + 3u_4)) - 8N^2(u_2 - u_4)^2 \\ d_4 &= N(2u_3 + 3u_4)^2[8(2u_3 + 3u_4) - 9(2N - 2(u_2 - u_4))] \\ e_4 &= (2u_3 + 3u_4)^3[2N - 2(u_2 - u_4) - (2u_3 + 3u_4)] , \end{aligned}$$

and

$$\begin{aligned} a_3 &= 11N^4 \\ b_3 &= 48N^4 - 11N^3(-3u_1 + u_3) \\ c_3 &= 36N^2(u_2 - u_4)^2 - 8N^2(3u_2 + 2u_3)(2u_3 + 3u_4) \\ d_3 &= -4N^3(3N - 3u_1 + u_3) \\ e_3 &= (3u_2 + 2u_3)^2(2u_3 + 3u_4)^2 . \end{aligned}$$

2.4 Quartic Del Pezzo surface with one \mathbb{A}_3 and two \mathbb{A}_1 type singular points.

Again consider the case of a reflexive polytope with four lattice points on the boundary, as in the graph below.



We once more we can see that the polytope generated by the lattice points $(1, 1), (0, 1), (-1, 1), (0, -1)$ in \mathbb{Z}^2 gives the same toric variety as the polytope generated by the lattice points $(1, 0, 3), (2, 2, 0), (1, 2, 1), (0, 2, 2)$ in \mathbb{Z}^3 .

The non-negative integer matrix

$$A = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 0 & 2 & 2 & 2 \\ 3 & 0 & 1 & 2 \end{bmatrix}$$

corresponds to a toric algebraic variety which is the topological closure $\overline{f(\mathbb{C}^3)}$ of the image $f(\mathbb{C}^3) \subset \mathbb{C}^4$ under the map

$$f : \mathbb{C}^3 \rightarrow \mathbb{C}^4, (\theta_1, \theta_2, \theta_3) \mapsto \frac{1}{(\theta_1 \theta_3^3 + \theta_1^2 \theta_2^2 + \theta_1 \theta_2^2 \theta_3 + \theta_2^2 \theta_3^2)} (\theta_1 \theta_3^3, \theta_1^2 \theta_2^2, \theta_1 \theta_2^2 \theta_3, \theta_2^2 \theta_3^2) .$$

According to Birch's theorem there is a unique maximum likelihood estimate $\hat{\theta}$ for the data $u = (u_1, u_2, u_3, u_4)$ which satisfies the equation

$$A \cdot \hat{p} = \frac{1}{N} \cdot A \cdot u ,$$

where $N = u_1 + u_2 + u_3 + u_4$ and \hat{p} is the probability distribution corresponding to the parameter values $(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3)$.

This gives us the equations

$$\begin{aligned} \hat{\theta}_1 \hat{\theta}_3^3 &= \frac{1}{N} u_1 \\ \hat{\theta}_1^2 \hat{\theta}_2^2 - \hat{\theta}_2^2 \hat{\theta}_3^2 &= \frac{1}{N} (u_2 - u_4) \\ \hat{\theta}_1 \hat{\theta}_2^2 \hat{\theta}_3 + 2 \hat{\theta}_2^2 \hat{\theta}_3^2 &= \frac{1}{N} (u_3 + 2u_4) . \end{aligned}$$

and these equations give us

$$\hat{p}_1 = \frac{u_1}{N}$$

We can compute the unique maximum likelihood estimate $\hat{\theta}$ for the data u , which is

$$\begin{aligned}\hat{\theta}_1 &= \sqrt[4]{\frac{u_1(u_2 - u_4)^3}{8N} \left(\sqrt[2]{\frac{(u_2 - u_4)^2 + 4(u_3 + 2u_2)(u_3 + 2u_4)}{(u_2 - u_4)^2}} - 1 \right)^3} \\ \hat{\theta}_2 &= \sqrt[4]{\frac{(u_2 + u_3 + u_4)(u_3 + 2u_4)^3}{Nu_1(u_2 + 2u_3 + 3u_4)^2}} \\ \hat{\theta}_3 &= \sqrt[4]{\frac{u_1(u_2 - u_4)}{2N(u_3 + 2u_4)} \left(1 + \sqrt[2]{\frac{(u_2 - u_4)^2 + 4(u_3 + 2u_2)(u_3 + 2u_4)}{(u_2 - u_4)^2}} \right)}\end{aligned}$$

2.5 Toric Del Pezzo surfaces of degree greater than five.

When the degree of the Del Pezzo surface is greater than five, the defining equations of the probability distribution $(\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4)$ have degree 5 and we cannot obtain a closed form for the Maximum Likelihood Estimate. Although a closed-form formula for maximum likelihood estimates is not achieved for these log-linear models, the log-likelihood function is convex for these models, and any hill-climbing algorithm can be used to compute the ML estimates.

References

- [1] F. Catanese, S. Hosten, A. Khetan and B. Sturmfels, *The Maximum likelihood degree*, American Journal of Mathematics, **18** (2006), 671-697.
- [2] M. Drton, B. Sturmfels, S. Sullivant, *Lectures on Algebraic Statistics*, Oberwolfach Seminars, **39** (2008), Birkhuser, Basel.
- [3] J. D. Hauenstein, J. Rodriguez, and B. Sturmfels, *Maximum likelihood for matrices with rank constraints*, Journal of Algebraic Statistics, **5** (2014), 18-38.
- [4] S. Hosten, A. Khetan and B. Sturmfels, *Solving the likelihood equations*, Foundations of Computational Mathematics, **5** (2005), 389-407.
- [5] M. Kreuzer, H. Skarke, *On the classification of reflexive polyhedra*, Commun. Math. Phys. 185 (1997) 495-508.
- [6] L. Pachter and B. Sturmfels, *Algebraic Statistics for Computational Biology*, (2005) Cambridge University Press, New York.
- [7] H. Sato, *Toward the classification of higher-dimensional Fano varieties*, Tohoku Math. J., **52** (2000), 383-413.

- [8] B. Sturmfels, Open problems in Algebraic Statistics, In "*Emerging Applications of Algebraic Geometry*", I.M.A. Volumes in Mathematics and its Applications, **149** (2009), Springer, New York, 351-364.

Dimitra Kosta
School of Mathematics,
The University of Edinburgh,
Edinburgh EH9 3JZ, UK.
D.Kosta@ed.ac.uk